



## VERİ MADENCİLİĞİ YÖNTEMLERİ İLE TARIMSAL VERİ ANALİZİ: BİR AKILLI TARIM SİSTEMİ ÖNERİSİ

Burak ÖZER<sup>1\*</sup>, Sümeyra KUŞ<sup>2</sup>, Oktay YILDIZ<sup>3</sup>

<sup>1</sup> Gazi Üniversitesi, Bilişim Enstitüsü, Bilişim Sistemleri, Ankara, Türkiye

<sup>2</sup> Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Ankara, Türkiye

<sup>3</sup> Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Ankara, Türkiye

### Anahtar Kelimeler

Akıllı Tarım,  
Veri Madenciliği,  
Veri Analizi,  
Coğrafi Bilgi Sistemleri.

### Öz

Gıda kıtlığı ve nüfus artışı, dünya çapında sürdürülebilir kalkınmanın karşı karşıya olduğu en büyük zorluklardır. Bununla beraber iklim değişikliğinin etkisi diğer birçok alanda olduğu gibi tarım alanında da belirsizliklere neden olacaktır. Yapay zekâ (AI), Nesnelerin İnterneti (IoT), Coğrafi Bilgi Sistemleri (CBS) ve mobil internet gibi gelişmiş teknolojiler, dünyanın karşı karşıya olduğu zorluklara gerçekçi çözümler sağlayabilmektedir. Günümüzde sensör cihazlarının yaygınlaşması ile tarım alanlarından veri elde etmek ve ham veriden bilgi üretmek yaygınlaşmıştır. Bu çalışma, Akıllı tarımda veri analizi ile ürün tahmini üzerine yoğunlaşmıştır. Kaggle üzerinden sağlanan toprak (ph, azot, potasyum ve fosfor değeri) ve iklim verileri (sıcaklık yağış ve nem) kullanılarak veri madenciliği algoritmaları ile farklı modeller oluşturulmuştur. Oluşturulan modeller doğruluk, kesinlik, duyarlılık, f-skor ve algoritmanın çalışma zamanına göre kıyaslanmıştır. Rastgele Orman algoritmasıyla geliştirilen model, çalışma süresi yaklaşık 0,05 s ve %99,5'lik doğruluk değeri ile en optimum sonuçları vermiştir. Daha sonra, Rastgele Orman algoritması Hindistan bakanlıklarınca sağlanan toprak verileri ve meteoroloji verilerine uygulanmış ve Hindistan'ın tarımsal ürün haritası oluşturulmuştur.

## AGRICULTURAL DATA ANALYSIS WITH DATA MINING METHODS: A SMART FARMING SYSTEM PROPOSAL

### Keywords

Smart Farming,  
Data Mining,  
Data Analysis,  
Geographic Information  
Systems.

### Abstract

Food scarcity and population growth are among the biggest challenges facing sustainable development worldwide. However, the impact of climate change will cause uncertainties in agriculture, as in many other areas. Advanced technologies such as artificial intelligence (AI), Internet of Things (IoT), Geographic Information Systems (GIS) and mobile internet can provide realistic solutions to the challenges facing the world. Today, with the widespread use of sensor device, it has become widespread to obtain and examine data from agricultural areas. This study focuses on product forecasting with data analysis in smart agriculture systems. Machine learning models are created using soil data (ph value, nitrogen value, potassium value and phosphorus value) and climate data (temperature, rainfall and humidity) provided via Kaggle. The created models are compared according to accuracy, precision, recall, f-score and the running time of the algorithm. The model developed with the Random Forest algorithm gave the most optimum results with a running time of approximately 0.05 s and an accuracy of 99.5%. Then, the Random Forest algorithm was applied to the data provided by the Indian ministries and finally, the agricultural product map of India is created.

### Alıntı / Cite

Ozer, B., Kus, S., Yildiz, O., (2022). Veri Madenciliği Yöntemleri ile Tarımsal Veri Analizi: Bir Akıllı Tarım Sistemi Önerisi, Mühendislik Bilimleri ve Tasarım Dergisi, 10(4), 1417-1429.

Yazar Kimliği / Author ID (ORCID Number)	Makale Süreci / Article Process	
B. Özer, 0000-0003-2615-9389	<b>Başvuru Tarihi / Submission Date</b>	04.03.2022
S. Kuş, 0000-0002-5288-769X	<b>Kabul Tarihi / Accepted Date</b>	27.04.2022
O. Yıldız, 0000-0001-9155-7426	<b>Yayın Tarihi / Published Date</b>	30.12.2022

## 1. Giriş (Introduction)

Tüm dünyada tarım sektörü milli gelirin önemli bir payını oluşturmaktadır. Bu nedenle, tarım sektörünü iyileştirmek için yeni teknolojilerin uygulanması, ulusal ekonomiyi desteklemek için önemli bir konudur. Tarımsal üretim, endüstriyel süreç için ihtiyaç duyulan hammaddelere ek olarak, insanlar ve hayvanlar için gıda üretimini içerir. Antik çağlardan bu yana, pek çok tarımsal kalkınma devrimi vardır; İlk tarım devrimi, eski insanların ilgisini tarım yöntemlerinin geliştirilmesinde yansıtan Mısır ve Yunan eski uygarlıkları tarafından yapılmıştır. İkinci tarım devrimi, Avrupa kıtasında feodalizmin sona ermesini takip eden 17. yüzyılda ortaya çıkmıştır. Üçüncü tarım devrimi (Yeşil Devrim) geçen yüzyılın 1930-1960 döneminde görülmüş; burada tarımsal üretimi artırmak için mineral gübrelerin kullanımının yaygınlaşması ve çeşitli tarım makinelerinin geliştirilmesine paralel olarak pestisit kullanımının artması (Bochtis) olmuştur. Dördüncü tarım devrimi, bilgi iletişim teknolojisi (BİT) ve yapay zekada önemli bir gelişme olan son yirmi yılda meydana gelmiştir. Bu teknolojiler, robotların hasat ve ot ayıklama gibi tarımsal işlemlerde kullanıldığı ekipman ve cihazların uzaktan kontrol edilmesini kolaylaştırmış ve ayrıca mahsulleri gübrelemek ve mahsulün büyüme aşamalarını izlemek için dronlar da kullanılmıştır (Mohamed vd., 2021).

Akıllı tarım, iklim faktörleri, toprak özellikleri, toprak nemi vb. değişikliklerin izlenmesine izin verdiği için yetiştirilecek ürün tahmini ve daha sonra eşzamanlı sağlanan veri ile diğer analizler gibi ilgili birçok konuyu ele almaktadır. Günümüzde IoT (Internet of Things), hemen hemen her alanda yaygınlaşmaya başlamıştır. Akıllı tarımda IoT sadece büyük çiftçilik uygulamalarında değil, aynı zamanda hayvancılığın bakımı, sera çiftçiliği, çiftliklerin yönetimi vb. alanlarda da kullanılabilir. IoT için kullanılan en önemli araç sensörlerdir. Sensörler, istenen analizi elde etmek için yorumlanan temel verileri toplayan cihazlardır. Tarım için, sensörler esas olarak azot, potasyum ve fosfor değerlerini ölçmek, hastalıkları ve topraktaki nem içeriğini tespit etmek için kullanılan okumaları almak için kullanılmaktadır. Akıllı tarım sistemlerinde IoT cihazları, yenilikçi uygulama araçları kullanılarak gerçek zamanlı olarak toplanan ve analiz edilen büyük miktarda değerli tarımsal veri üretir (Ratnaparkhi vd., 2020; Roukha vd., 2020).

Yaşanan teknolojik gelişmelere rağmen iklim değişikliği, tarımda yaşanan en büyük zorluklardan birisidir. İklim değişikliğinin olumsuz etkileri, artan sıcaklıklar, hava değişkenliği, değişen tarımsal ekosistem sınırları, istilacı mahsuller ve zararlılar ve daha sık görülen aşırı hava olayları şeklinde şimdiden hissedilmektedir. Çiftliklerde iklim değişikliği mahsul verimini, başlıca tarım besin kalitesini düşürmektedir. Mevcut verimi korumak ve talebi karşılamak için üretim ve gıda kalitesi artışlarını sağlamak için adaptasyona önemli yatırımlar gerekecektir (CLIMATE-SMART AGRICULTURE, 2021). İklim değişikliği ile belirli bir bölgede yetişen tarım ürünleri de adaptasyon sürecine girmektedir. Belirli tarım ürünleri değişen iklim koşullarıyla beraber artık o bölgede yetişemeyecek duruma gelmektedir. Tarımla uğraşan insanların buna karşı tedbir almaları kaçınılmazdır. Mevcutta çoğu arazide yanlış ürün tercihleri yapılmaktadır. Bazı bölgelerde yıl içinde iklim değerlerine göre iki kez hasat yapılacaktır yanlış tarım ürünleri yetiştirme tercihinden dolayı bir kez hasat yapılabilir. Ayrıca çoğu bölgelerde yetişen tarım ürünleri o bölgenin iklimine duyarlı olduğu için tarım ürününden yeterince verim alınmamaktadır.

Bu makalede ürün yetiştiricileri için bölgesel toprak özelliklerinin yanı sıra iklim şartlarını da dikkate alarak ürün önerisinde bulunan akıllı tarım sistemi geliştirilmiştir. Tarımı etkileyen parametreler arasındaki gizli kalıpları çıkarmada veri madenciliğinden faydalanılmıştır. Oluşturulan ürün haritası, bir bölgede yetişecek doğru tarım ürünlerini göstermektedir. Diğer bir deyişle, o bölgedeki sıcaklık, nem, yağış ve toprak özelliklerine göre yetiştirilecek en uygun tarım ürünü belirlenmektedir. Oluşturulan ürün haritası, tarımla uğraşan insanlara yaşadıkları bölgenin özelliklerine göre yetişebilecek en doğru tarım ürünlerini belirlemede bir rehber niteliğinde olacaktır. Ayrıca tarım bir ekonomik değerdir ve oluşturulan tarımsal ürün haritası sayesinde, ekonomisinin büyük bölümü tarım üzerine olan ülkelerde doğru tarım ürünlerinin tercihi ile elde edilen tarımsal üretim artmaktadır. Dolayısıyla tarım ülkelerinin ekonomilerine büyük katkı sağlanması hedeflenmektedir.

## 2. Literatür taraması (Literature review)

Makalenin bu bölümünde literatürdeki benzer çalışmalar incelenmiştir.

Tarımsal bitkilerin optimum şekilde sulanmasını sağlayacak bir sistem, akıllı telefon ve web uygulaması kullanılarak mahsul alanında kullanılan sensörlerden alınan verinin analizini yapacak şekilde geliştirilmiştir.

Birliktelik kuralları (Association rules), bu çalışmada uygulanan veri madenciliği tekniğidir. Sıcaklık, nem ve toprak nemi ile limon verimleri ve evde yetiştirilen sebze arasındaki ilişkileri bulmak için kullanılmışlardır. Apriori algoritması bu çalışmada kullanılan birliktelik kuralları modelidir. Ayrıca, birkaç girdi değişkeni ile bir çıktı değişkeni arasındaki ilişkiyi modellemek için doğrusal regresyon modeli uygulanmıştır. Sonuçlar tarıma açık faydalar göstermiştir. Topraktaki nem içeriği sebzeler için uygun şekilde kontrol edilmiş, böylece maliyetler düşürülmüş ve tarımsal verimlilik artmıştır (Muangprathuba vd., 2018).

İklim değişikliklerine göre hasat zamanını planlamak, geçmiş çiftlik verilerine dayalı mahsul hastalıklarını önleyici olarak azaltmak ve yıllık mahsul üretimini tahmin etmek gibi görevlerde çiftçilere yardımcı olmak için örnek bir Kolombiya çiftliğinde uygulanan IoT-Agro sistemi geliştirilmiştir. Bu çalışma diğer çalışmaların eksikliklerini görerek üç katmanlı bir mimariye (Tarımsal Algı, Sınır Bilişim (Edge Computing) ve Veri Analitiği) dayalı bir Akıllı Tarım Sistemi önermiştir. Veri Analitiği Katmanında, Karar Ağaçları, Yapay Sinir Ağları, XGBoost, Destek Vektör Makinesi ve Rastgele Orman gibi kahve üretimini tahmin etmek için farklı algoritmalar değerlendirildi. XGBoost algoritması en düşük Ortalama Mutlak Hatayı (0.032) elde etmiştir (Rodríguez vd., 2021).

Mahsul hasadı tahmininden eksik veya yanlış sensör verilerinin yeniden yapılandırılmasına, çeşitli makine öğrenimi tekniklerinden yararlanma ve karşılaştırmaya kadar çeşitli pratik görevlerin tasarımı ve konuşlandırılması mümkün hale gelmiştir. Bu çalışmaya göre, İstat veri setindeki elma ve armut toplam mahsullerinin tahminine %90'a yakın bir başarı oranlarına sahip bir sinir ağı modeli ile ulaşılabileceği gösterilmiş, CNR bilimsel verileri için polinomsal regresyon ve karar ağaçlarının eksik verilerin tahmininde kullanışlı olduğu belirlenmiştir (Balducci vd., 2018).

Mahsul ekimleri arasındaki herhangi bir zamanda, bitki hastalıkları yaprağı etkileyebilir ve bu da mahsul üretiminde büyük zararlara ve ekonomik kayıplara neden olabilmektedir. Bu çalışmada, Veri Madenciliği algoritmaları (Destek Vektör Makinesi, Rastgele Orman ve Olasılıksal Dereceli Azalma) ve Derin Öğrenme (Inception-v3, VGG-16, VGG) algoritmalarının turuncğiller bitkisi hastalık tespiti açısından performansları karşılaştırılmıştır. Hastalık sınıflandırma doğruluğu (CA), hastalık tespiti durumunda Derin Öğrenme yöntemlerinin Makine Öğrenmesi yöntemlerinden daha iyi performans göstermiştir. VGG-16, CA açısından en iyi değerleri göstermiştir (Sujatha vd., 2021). Zararlı böcekler, tüm dünyada tarımsal ürün verimini ve kalitesini etkileyen başlıca nedenlerden biridir. Yapılan çalışmanın amacı araştırmacılara ve teknisyenlere, Çeşitli akıllı haşere izleme (SPM) modellerinin uygulanmasını teşvik edebilecek Derin Öğrenme tekniklerini ve akıllı haşere izlemede kullanılan son teknolojileri özetlemektir. Bu çalışmaya göre böcek zararlılarının sınıflandırma yöntemleri üç açıdan özetlenebilir: Klasik CNN ağlarını geliştirmek, transfer öğrenme ve birkaç adımlı öğrenme (Li vd., 2021).

IoT ve akıllı görüntü tanıma dayalı bir hasat sisteminde sinir ağı modelleri eğitilerek nesne tespiti yoluyla mahsul olgunluğu belirlenebilir ve daha sonra olgun mahsuller robotik kollar kullanılarak hasat edilebilmektedir. Ampirik kanıtlar, önerilen MobileNet-SSD model eğitiminin diğer modellerden daha yüksek olan %84'lük bir ortalama kesinliğe (mAP) sahip olduğunu göstermiştir. Araştırma yöntemi olarak MobileNet-SSD modeli seçilmiştir (Horng vd., 2019). Bina çatıları, açık bahçeler ve kapalı tarım gibi kentsel alanlarda çiftçilik yapılabilir hale gelmiştir. Bu çalışmada, kentsel tarım bağlamında nem, sıcaklık ve toprak nemini gerekli tarım parametreleri olarak kabul eden IoT tabanlı bir Smart AgroTech sistemi önerilmiştir. Önerilen sistem, tarım arazisinin durumuna bağlı olarak sulama eyleminin başlayıp başlamayacağına karar verir ve çiftlik sahibine izleme olanağı ve uzaktan kontrol sağlar. Önerilen denklem, nem ve toprak nemi için ortalama hata oranını %3'ün altında ve sıcaklık için %1,5'in altında değer bulmuştur (Podder vd., 2021).

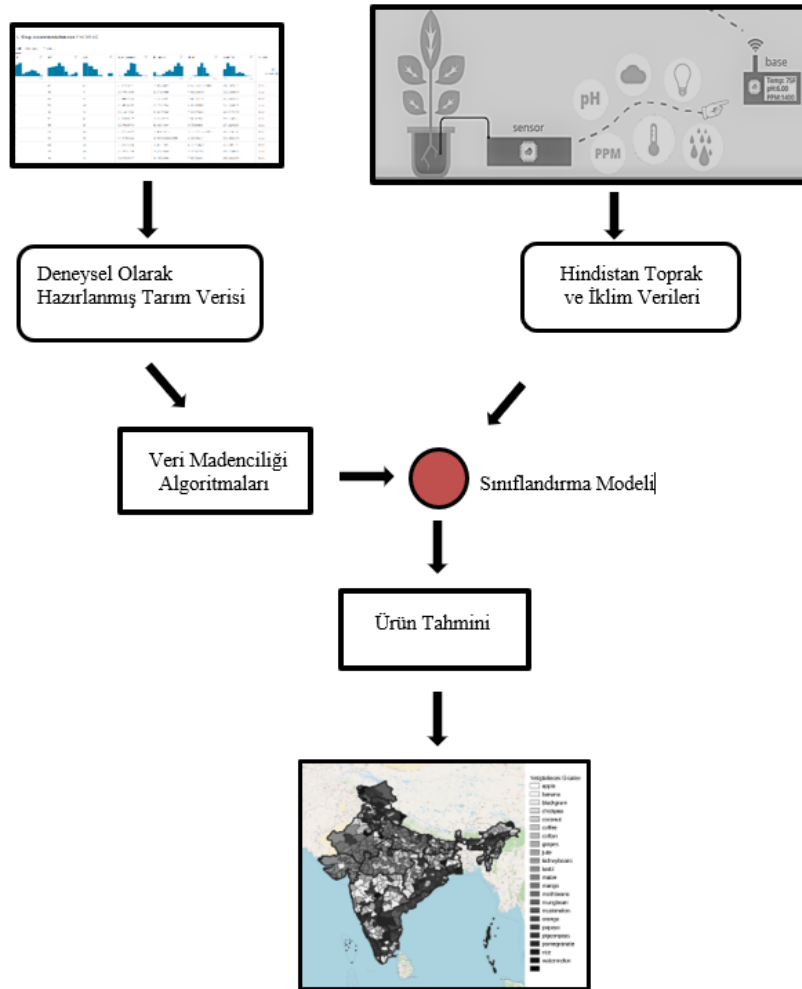
Guguk Kuşu Arama Algoritması ile suyun her koşulda çiftçilik için tahsis edilmesi sağlanmaktadır. Sıcaklık, bulanıklık, pH, nem gibi çeşitli veriler, ilgili sensörler ve kablosuz iletişim sistemleri ile donatılmış Nesnelerin İnterneti (IoT) platformu kullanılarak toplanmıştır (Pathak vd., 2019). Çok Katmanlı Algılayıcı (MLP) yaklaşımı, IoT tarımında şeker verimi üretim miktarını tahmin etmek için geliştirilmiştir. Deneysel sonuçlar, önerilen MLP algoritmasının şeker kamışı verimini tespit etmek için maksimum %99 doğruluğa, %95 hassasiyete, %96 duyarlılık ve %0.04 Minimum Ortalama Mutlak Hataya (MAE) ve %0.006 Ortalama Karekök Hatasına (RMSE) sahip olduğunu göstermektedir (Wang vd., 2021).

Literatürdeki çalışmalar incelendiğinde akıllı tarım sistemlerinde veri analizinin genellikle iki ana bölümde yapıldığı gözlemlenmiştir. Yapılan çalışmalar (Muangprathuba vd., 2018; Rodríguez vd., 2021; Balducci vd., 2018; Podder vd., 2021; Pathak vd., 2019; Wang vd., 2021) sensor, gps vs. cihazlardan elde edilen yapısal özellikteki toprak ve iklim verilerinin veri madenciliği algoritmaları kullanılarak analiz edilmesinden oluşmaktadır. Sensor verileri toprağın özellikleri (azot değeri, potasyum değeri vs.), iklim verilerinden oluşmaktadır. Diğer çalışmalarda (Sujatha vd., 2021; Li vd., 2021; Horng vd., 2019) derin öğrenme ile görüntü sınıflandırma ve görüntü algılama ile tarım alanlarından ve tarım ürünlerinden elde edilen yapısal olmayan görüntü verilerinin analiz edilmesinden oluşmaktadır.

Bu makalede yapısal özellikli veriler kullanılmıştır. Kullanılan veri, iklim (sıcaklık, yağış, nem) ve toprağın azot, fosfor, potasyum ve pH değerlerinden oluşmaktadır. Daha sonra, veri madenciliği algoritmalarından olan K-en yakın komşu, Rastgele Orman ve XGBOOST algoritmaları kullanılarak model oluşturulmuştur. Hindistan bölgesinden elde edilen iklim ve toprak verileri sayısallaştırılmış ve en yüksek başarıyı veren model ile ürün haritası oluşturulmuştur.

### 3. Materyal ve Yöntem (Material and Method)

Makalenin bu bölümünde, önerilen sistemin metodolojisi ve çalışmada kullanılan veri kümesine ilişkin bilgilere yer verilmiştir. Makalede önerilen sistemin genel yapısı Şekil 1’de özetlenmiştir. Buna göre önerilen Akıllı Tarım sisteminde veri analizi, deneysel olarak hazırlanmış tarım verisinin veri madenciliği algoritmaları kullanılarak bir sınıflandırma modeli oluşturulur. Daha sonra Hindistan toprak ve iklim verileri oluşturulan sınıflandırma modeline göre düzenlenir ve ürün tahmini için kullanılır. Çalışmanın son kısmında Hindistan için ürün haritası oluşturulmuştur.



Şekil 1. Sistem metodolojisi (System methodology)

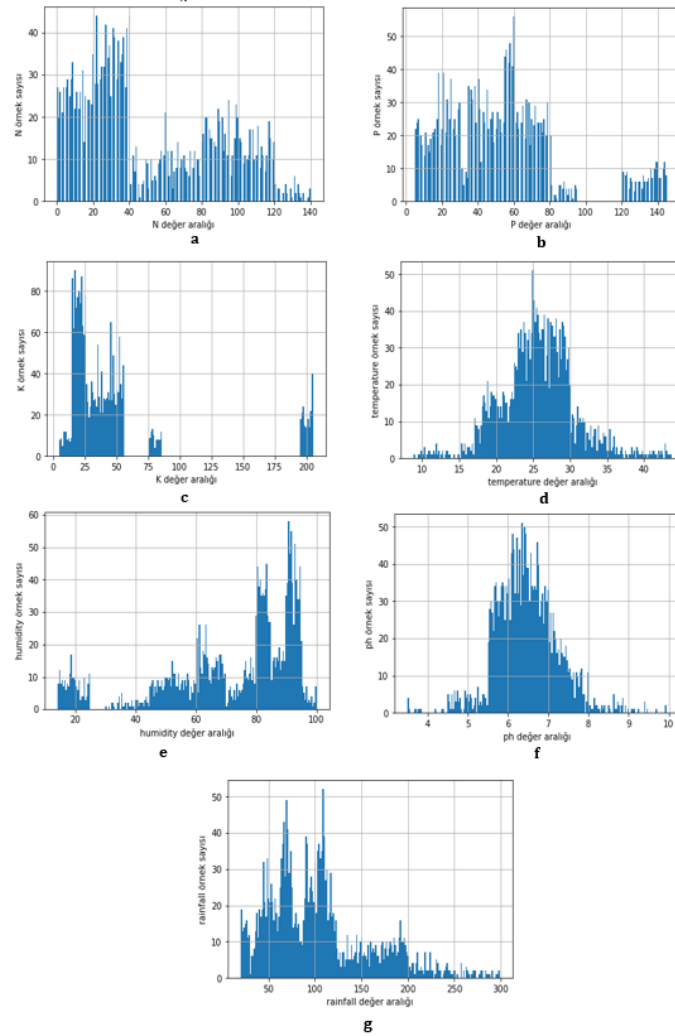
#### 3.1 Veri kümesi

Makalede kullanılan veri kümesi iki bölümden oluşmaktadır. Model oluşturmada kullanılan veri kümesi (Ingle, 2020) Kaggle üzerinden sağlanmıştır. Bu veri kümesi, Hindistan için mevcut olan yağış, iklim ve gübre verilerinin veri kümelerini artırarak oluşturulmuştur. Veri kümesi kullanıcıların çeşitli parametrelere dayalı olarak belirli bir çiftlikte yetiştirmek için en uygun mahsulleri önermek için tahmine dayalı bir model oluşturmalarını sağlar.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

Şekil 2. Kaggle tarımsal veri kümesi (Kaggle agricultural dataset)

Makalede kullanılan veri kümesi Şekil 2’de gösterilmiştir Şekilde görüldüğü gibi veri kümesi 2200 örnekten ve 7’si bağımsız 1’i bağımlı olmak üzere 8 özelliğten oluşmaktadır. Bu veri kümesindeki özellikler topraktaki Azot içeriği oranı (N), topraktaki Fosfor içeriği oranı (P), topraktaki Potasyum içeriği oranı (K), santigrat derece cinsinden sıcaklık değeri (temperature), yüzde olarak bağıl nem (humidity), toprağın ph değeri (ph), mm cinsinden yağış (rainfall) ve bu 7 özelliğe bağlı olarak değişen Hindistan’da yetişen 22 farklı bitki türünü ifade eden etikettir (label).



Şekil 3. a) Azot değer aralığı, b) Fosfor değer aralığı, c) Potasyum değer aralığı, d) Sıcaklık değer aralığı, e) Nem değer aralığı, f) Ph değer aralığı, g) Yağış değer aralığı (a) Nitrogen value range, b) Phosphorus value range, c) Potassium value range, d) Temperature value range, e) Humidity value range, f) Ph value range, g) Precipitation value range)

Kullanılan veri kümesinde değişkenlerin histogram grafikleri Şekil 3'te ayrı ayrı gösterilmiştir. Bu grafiklerden daha sonra modelde kullanılacak veri kümesinin değer aralıklarını belirlemede yararlanılmıştır. Hindistan bakanlıklarından sağlanan toprak verileri (ph, azot, potasyum, fosfor) ve iklim verileri (sıcaklık, nem, yağış) bu doğrultuda oluşturulmuştur. Bu veriler QGIS ortamında sayısallaştırılarak oluşturulan modele göre parametre değerleri ayarlanmıştır.

Etiket değişkeni Hindistan da yetişen farklı ürün çeşitlerini içermektedir. Bunlar, pirinç, mısır, nohut, barbunya, güvercin bezelyesi, güve fasulyesi, maş fasulyesi, siyah mercimek, yeşil mercimek, nar, muz, mango, üzüm, karpuz, kavun, elma, portakal, papaya, hindistan cevizi, pamuk, hint keneviri ve kahvedir.

### 3.2 Veri madenciliği yöntemleri

Deneysel çalışma kapsamında, veri madenciliğinde sıklıkla kullanılan sınıflandırma algoritmalarından K-En Yakın Komşu, Rastgele Orman ve XGBoost algoritmaları kullanılmıştır.

#### 3.2.1 K-en yakın komşu (KNN)

Komşu tabanlı bir sınıflandırma olan KNN, örnek tabanlı öğrenmenin veya genelleştirmeyen öğrenmenin bir türüdür. KNN, genel bir dahili model oluşturmaya çalışmaz, yalnızca eğitim verilerinin örneklerini depolar. Sınıflandırma, her noktanın en yakın komşularının basit çoğunluk oyu ile hesaplanır: bir sorgu noktasına, noktanın en yakın komşuları içinde en fazla temsilciye sahip olan veri sınıfı atanır (Nearest Neighbors, 2022).

```

Sınıflandır (X, Y, x) // X: eğitim verisi, Y: X sınıf etiketleri, x: bilinmeyen örnek
for i = 1 to m do
    Hesapla mesafe  $d(X_i, x)$ 
end for
k en küçük uzaklık  $d(X_i, x)$  için indisleri içeren I hesap kümesi.
return  $\{Y_i \mid i \in I\}$  için çoğunluk etiketi

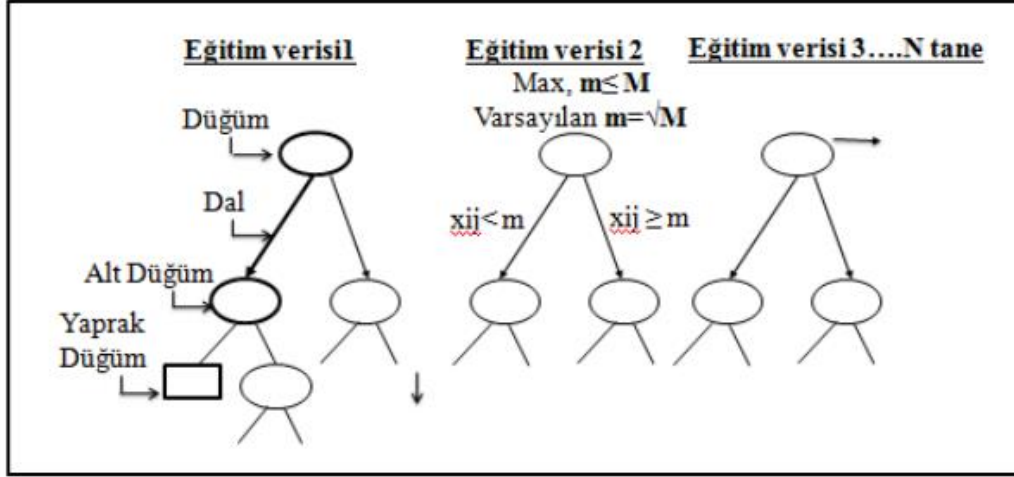
```

Şekil 4. KNN sınıflandırması için sözde kod (Pseudocode for KNN classification) (Tay vd., 2014)

Şekil 4'te KNN algoritmasının genel işleyişi özetlenmektedir. KNN algoritmasının performansı için kritik öneme sahip noktalardan birisi örnekler arası yakınlığın nasıl ölçümleneceğidir. Yakınlık, Öklid uzaklığı ya da bir başka uzaklık ölçütü kullanılarak hesaplanabilir. Algoritma 1'den de görülebildiği gibi, KNN algoritması basit bir yapıya sahiptir ve az sayıda parametre gerektirmektedir. Parametre seçiminde farklı teknikler kullanılmaktadır. Grid Search bu çalışmada parametre seçiminde kullanılmıştır.

#### 3.2.2 Rastgele orman

Rastgele Orman algoritması, veri kümesinin çeşitli alt örneklerinde bir dizi karar ağacı sınıflandırıcısına uyan bir meta tahmin edicidir ve tahmin doğruluğunu iyileştirmek ve aşırı öğrenmeyi kontrol etmek için karar ağaçlarının ortalamasını kullanır. Burada ağaçlar rastgele seçilmektedir. Bu rastgelelik kaynağının amacı, Rastgele Orman tahmincisinin varyansını azaltmaktır. Karar ağaçları tipik olarak yüksek varyans sergiler ve fazla uyum gösterme eğilimindedir. Rastgele Orman'da rastgelelik, bir şekilde ayrılmış tahmin hatalarına sahip karar ağaçları verir. Bu tahminlerin bir ortalaması alınarak bazı hatalar ortadan kaldırabilir. Rastgele Orman, çeşitli ağaçları birleştirerek, bazen sapmada hafif bir artış pahasına, azaltılmış bir varyans elde eder. Uygulamada, varyans azalması genellikle önemlidir, dolayısıyla genel olarak daha iyi bir model vermektedir (Ensemble methods, 2022)



Şekil 5. Rastgele Orman yöntemine ait ağaç yapısı (Tree structure of the Random Forest method) (Ok vd., 2011)

Şekil 5'te  $x_{ij}$ , girdi verilerini temsil etmektedir.  $M$  toplam değişken sayısını,  $m$  ise rastgele seçilir.  $m$  değerinin  $M$  (toplam değişken sayısı) değişkeninin kareköküne eşit olarak alınması genellikle optimum sonuçlar vermektedir.

Rastgele Orman algoritmasında ana parametreler  $N$  ( $n\_estimators$ ) ve  $M$  ( $max\_features$ )'dir.  $N$  ( $n\_estimators$ ) ormandaki ağaç sayısıdır. Ne kadar büyük olursa o kadar iyidir, ancak hesaplamak o kadar uzun sürmektedir.  $M$  ( $max\_features$ ), bir düğümü bölerken dikkate alınması gereken rastgele özellik alt kümelerinin boyutudur. Ne kadar düşük olursa, varyansın azalması o kadar büyük olur, fakat aynı zamanda yanlılıktaki artış da o kadar büyük olmaktadır.

### 3.2.3 XGBOOST algoritması

XGBoost, yüksek verimli, esnek ve taşınabilir olacak şekilde tasarlanmış optimize edilmiş bir dağıtılmış gradyan artırma tekniğidir. XGBoost hem sistem (donanım ve yazılım) optimizasyonu hem de algoritmik geliştirmeler ile günümüzde yaygın olarak kullanılmaktadır.

XGBoost Sistem Optimizasyonları:

- Paralel Çalıştırma: XGBoost, karar ağaçlarını oluştururken paralelizasyon yaparak çok daha hızlı oluşturulmasını sağlar. Bunu yapabilmesinin altında temel-öğrenenleri (base-learners) oluştururken, iç ve dış döngüler arasında geçiş yapabiliyor olması vardır. Normalde dış döngüler, karar-ağacının yapraklarını oluştururken iç döngüler de öznitelikler hesaplanır. Ancak, iç döngüler bitmeden dış döngüler tamamlanamayacağı için yani öznitelikler hesaplanmadan ağacın yaprakları oluşmayacağı için paralelizasyon sınırlanır. XGBoost, iç ve dış döngülere ayrılan hesaplama gücünü değiştirerek çalışma zamanını hızlandırır.
- Ağaç-Budama: XGBoost, en başından maksimum derinlik( $max\_depth$ ) parametresi ile ağacın derinliğini belirleyerek, eğer ağaç aşağı yönde fazla ilerledi ise geriye doğru budama yapar. XGBoost derinliğe öncelik verdiği için karmaşıklığı, dolayısıyla hesaplama performansını önemli ölçüde artırır.
- Donanım Optimizasyonu: XGBoost, en başta geliştirilirken donanım kaynaklarını daha iyi kullanmak üzere tasarlanmıştır. Bunun dışında, "out-of-core" hesaplama gibi iyileştirmeler sayesinde disk alanını optimize ederek daha büyük verileri belleğe sığdırabilmektedir.

XGBoost Algoritmik İyileştirmeler:

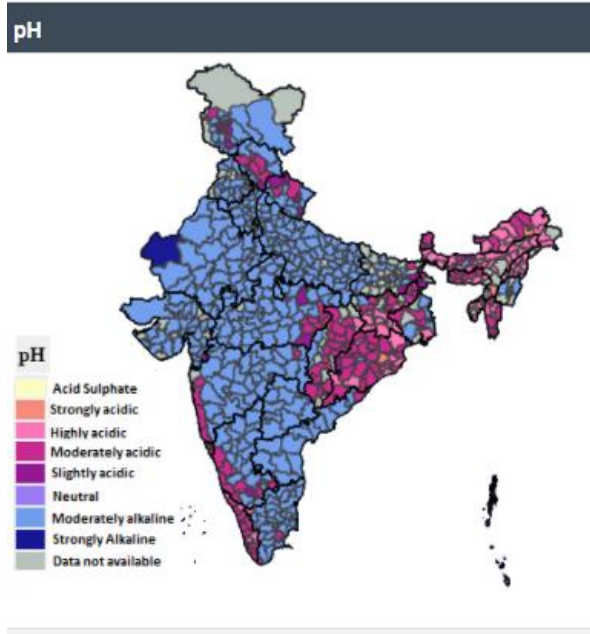
- Regülasyon: Hem LASSO hem de Ridge regularizasyonu kullanarak aşırı öğrenme engellenir.
- Seyreklik Uyumu: Gerçek hayatta veri setleri maalesef birçok eksik değer bulundurur. XGBoost, zayıf öğrenenler ile kayıp değeri eğitim kaybına bakarak en doğru şekilde öğrenebilmektedir.
- Ağırlıklı Çeyrek Çizim: XGBoost'un en büyük avantajlarından biri ağaçlara ayırırken en doğru noktadan ayırabilmek için veri setindeki gözlem noktalarını ağırlandırarak kullanıyor olmasıdır.
- Çapraz-doğrulama: XGBoost, kendi içinde cross-validation ( $cv$ ) uygulamasını yapabilmektedir. Ayrıca her çalıştırmada kaç iterasyon yapılacağını belirtmesine gerek yoktur (XGBoost Documentation, 2022).



#### 4. Deneysel çalışma (Experimental study)

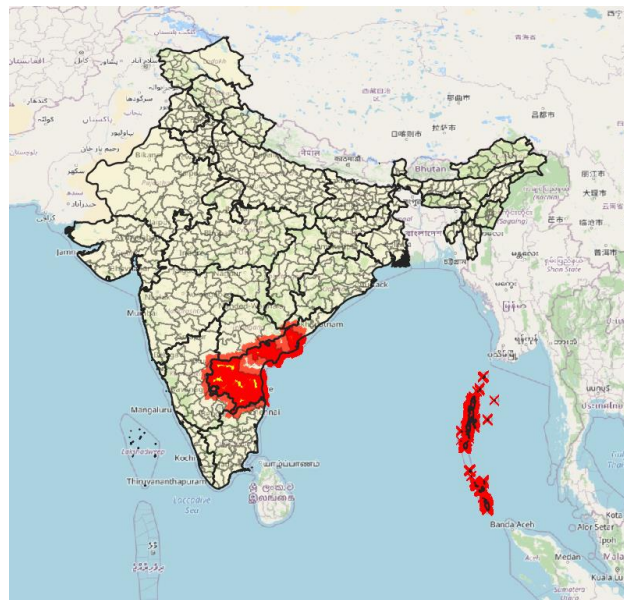
Bu çalışmada veri madenciliği yöntemi olan sınıflandırma algoritmaları kullanılmıştır. Sınıflandırma, verilerdeki kalıpları belirlemek için algoritmaları kullanır. Tahmin yapabilen bir veri modeli oluşturmak için de bu kalıplar kullanılır (What is machine learning?, 2021).

Bu çalışmada kullanılan sınıflandırma algoritmaları KNN, Rastgele Orman, XGBOOST sınıflandırma algoritmalarıdır. Çalışma Jupyter Notebook ortamında modellenmiştir. TensorFlow, Keras, Scikit-learn, pandas, numpy, matplotlib kütüphanelerinden faydalanılmıştır. Sınıflandırma algoritmaları, çalışma sürelerine ve doğruluk, kesinlik, duyarlılık ve F1 skor metriklerine göre iki farklı şekilde değerlendirilmiştir.



Şekil 6. Hindistan haritası ph değerleri (India map ph values)

Çalışmanın ikinci bölümünde Hindistan bakanlıklarınca yayınlanan toprak verileri (Soil Maps- Cycle I, tarih yok) ve iklim verileri (Annual Rainfall Map of India, 2021), (India - Current Temperature [°C], 2021), (Average Humidity for India in January, tarih yok), (South Asia Network on Dams, Rivers and People, tarih yok) Qgis ortamında sayısallaştırılmıştır. Şekil 4, Hindistan topraklarının ph değerlerini göstermektedir. Buna göre lejant aralığından ph değerleri Qgis ortamında sayısallaştırılmıştır.



Şekil 7. Qgis ortamında oluşturulmuş Hindistan haritası (India map created in Qgis environment)

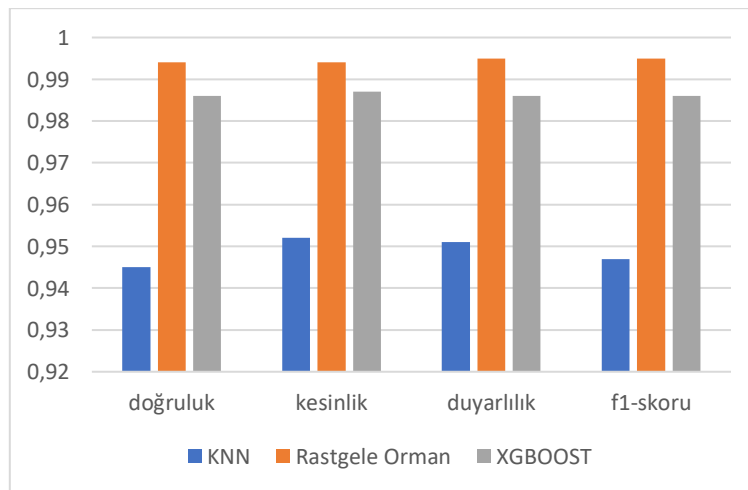


	NAME_1	NAME_2	ph	Nitrogen	Phosphorus	Potassium	first_temp	sec_temp	first_rain	sec_rain	first_hum	s
1	Andaman and ...	Andaman Islands	5,200	45	27	29	29	25	301	1503	75	
2	Andaman and ...	Nicobar Islands	4,500	46	26	22	28	24	301	1503	75	
3	Andhra Pradesh	Anantapur	4,700	13	25	35	26	27	59	492	74	
4	Andhra Pradesh	Chittoor	5,300	34	24	5	32	25	59	492	70	
5	Andhra Pradesh	Cuddapah	5,400	38	25	30	27	28	59	492	70	
6	Andhra Pradesh	East Godavari	5,300	35	25	20	31	24	59	492	70	
7	Andhra Pradesh	Guntur	6,300	32	24	20	31	26	59	492	70	
8	Andhra Pradesh	Krishna	5,400	32	91	23	33	28	59	492	70	
9	Andhra Pradesh	Kurmoor	6,400	26	49	35	30	25	59	492	70	
10	Andhra Pradesh	Nellore	6,300	21	63	20	31	26	59	492	70	

**Şekil 8.** Qgis ortamında sayısallaştırılmış ve tablolaştırılmış veri kümesi (Digitized and tabulated dataset in Qgis environment)

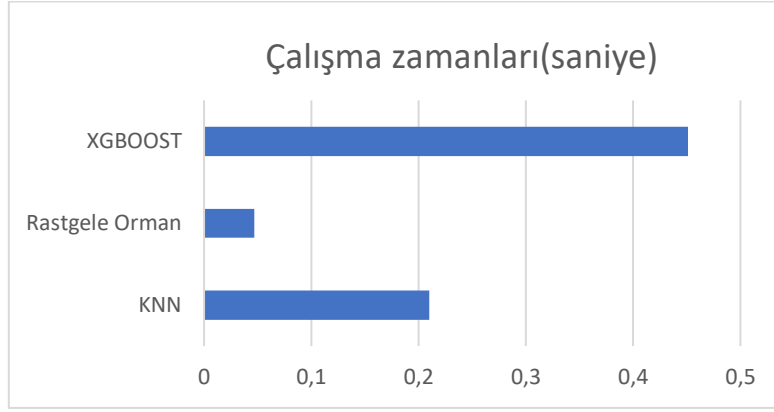
Şekil 7’de Hindistan haritası ve örnek olarak seçilen 10 nokta gösterilmiştir. Şekil 8’de ise bu haritanın tablolaştırılmış hali gösterilmektedir. Şekil 8’de NAME\_1 Hindistan’da bulunan eyaletleri NAME\_2 Hindistan’da bulunan şehirleri ve diğer değişkenler Hindistan bakanlıklarınca sağlanan haritalardan yola çıkılarak oluşturulmuştur. İklim verileri (temp, rain, hum) yine aynı şekilde bazı haritalardan ve meteoroloji verilerinden elde edilerek oluşturulmuştur. Burada “first” ayın ilk altı ayının ortalama iklim verilerini belirtirken “sec” terimi ikinci altı ayı belirtmektedir. Örneğin “first\_temp” Hindistan’ da 2019 yılının ilk altı ayının sıcaklık ortalamasının gösterirken “sec\_temp” ise ikinci altı ayın sıcaklık ortalamalarını göstermektedir.

Gerçekleştirilen çalışmada sınıflandırma algoritmasında en uygun parametre değerlerinin seçimi için Izgara Araması (Grid Search) kullanılmıştır. Izgara aramasında KNN için n\_neighbors (komşu sayısı) 2-15 aralığındaki tam sayılar denenmiş, Rastgele Orman için n\_estimators (ağaç sayısı), 5 ile 50 arasında 5’er artacak şekilde ve max\_depth (maksimum derinlik), 10 ile 50 arasında 10’ar artacak şekilde değerler ve XGBOOST için n\_estimators (ağaç sayısı), 5 ile 50 arasında 5’er artacak şekilde, max\_depth (maksimum derinlik), 10 ile 50 arasında 10’ar artacak şekilde ve learning\_rate (öğrenme oranı), 0.01, 0.1, 1 değerlerinin farklı bütün kombinasyonları denenmiş ve en yüksek doğruluk değerini veren kombinasyon her algoritma için seçilmiştir. KNN için n\_neighbors= 3, Rastgele Orman için n\_estimators=20, max\_depth= 45 ve XGBOOST için max\_depth = 10, n\_estimators = 20, learning\_rate = 0.1 parametre değerleri en yüksek doğruluk oranlarını vermiştir.



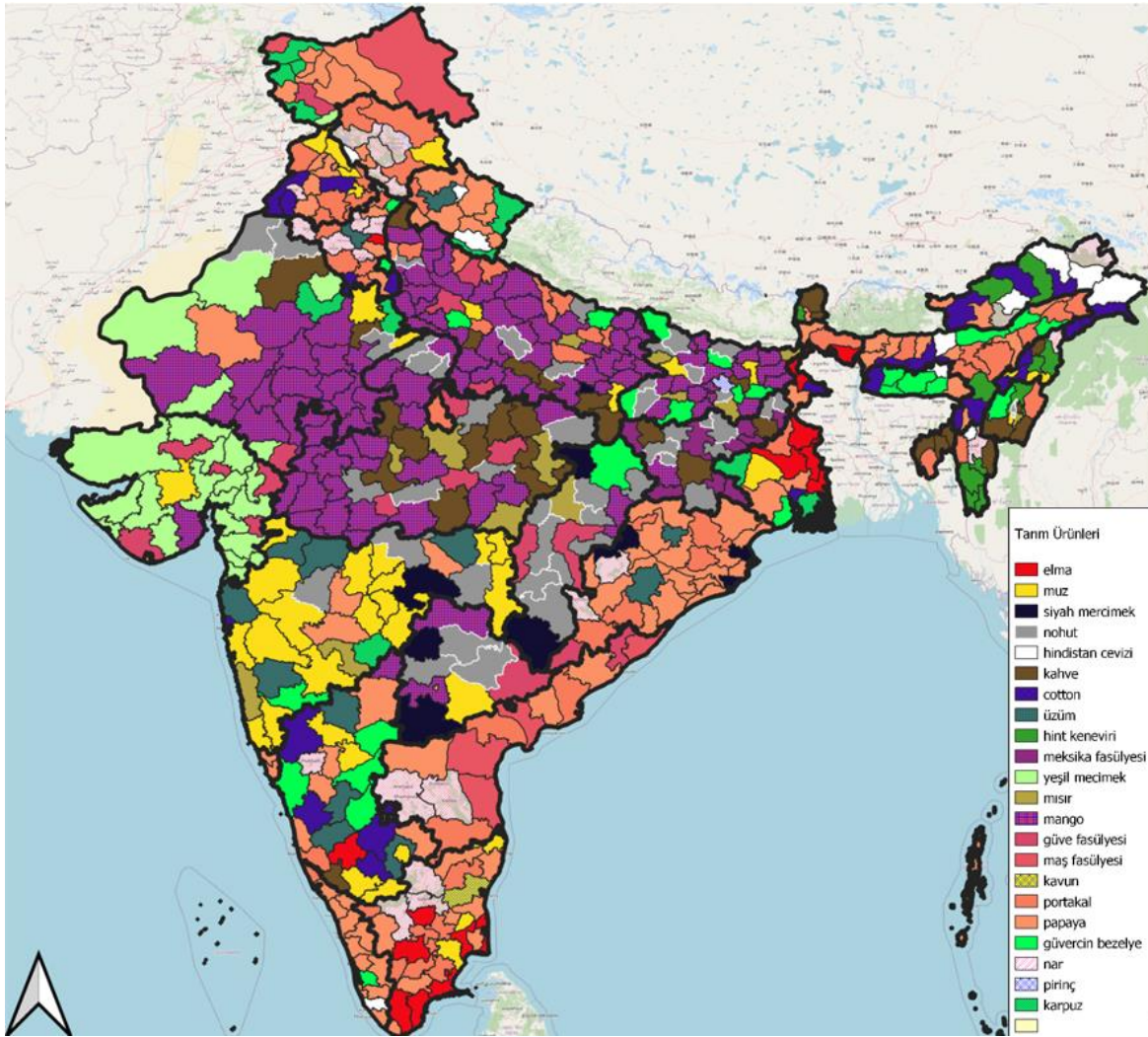
**Şekil 9.** Sınıflandırma algoritmalarının doğruluk, kesinlik, duyarlılık ve f1- skoru değerlerine göre karşılaştırılması (Comparison of classification algorithms according to accuracy, precision, sensitivity and f1-score values)

Şekil 9’da sınıflandırma algoritmaları doğruluk, kesinlik, duyarlılık ve f1- skoru değerlerine göre karşılaştırılmıştır. Şekilde de görüldüğü gibi Rastgele Orman algoritması diğer algoritmalara göre daha iyi değerler vermiştir.

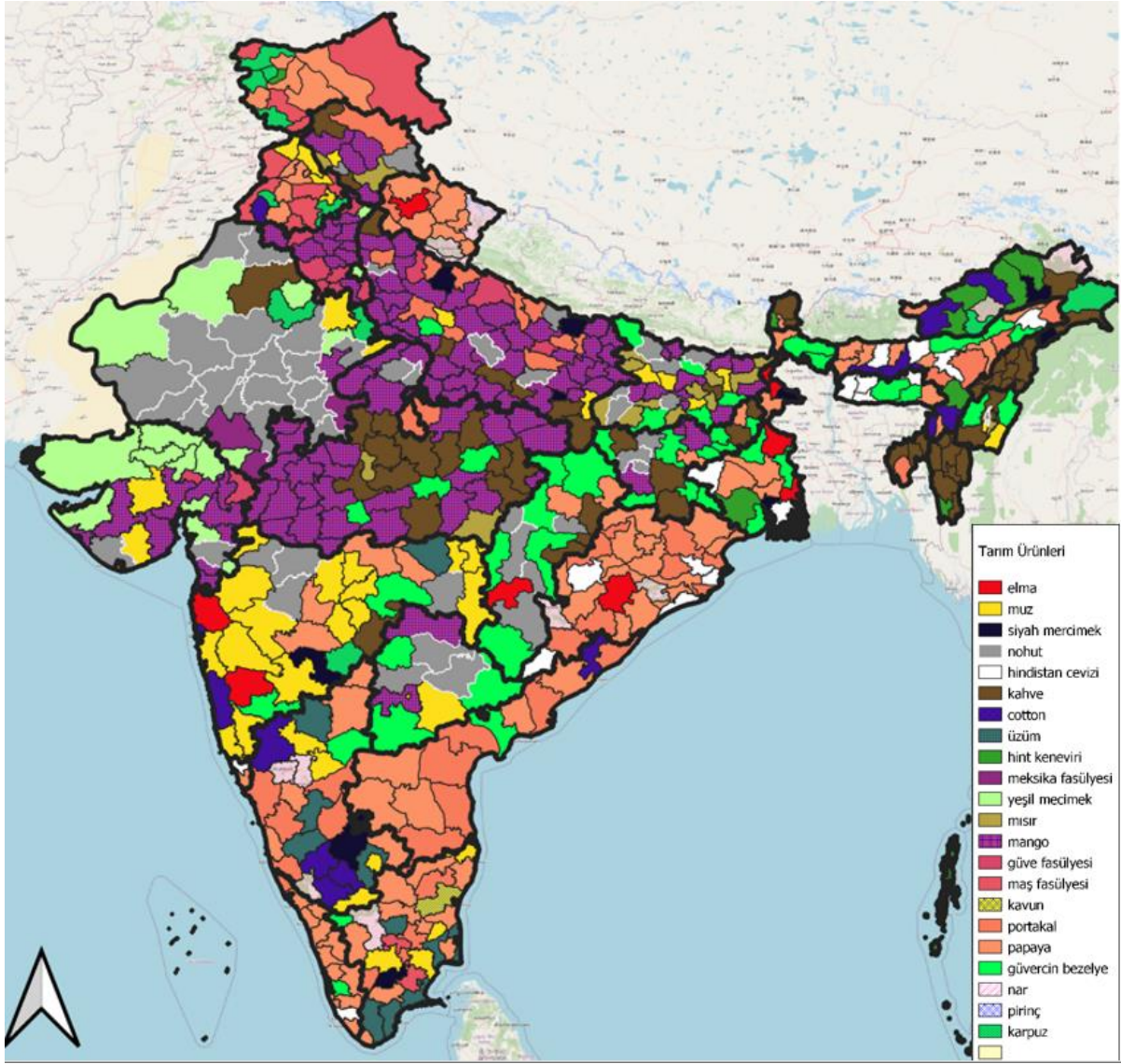


**Şekil 10.** Sınıflandırma algoritmalarının çalışma zamanlarına göre karşılaştırılması (Comparison of classification algorithms by runtime)

Şekil 10'da sınıflandırma algoritmaları çalışma zamanlarına göre karşılaştırılmıştır. Şekil 10'a göre Rastgele Orman çalışma zamanı diğer algoritmalara göre daha kısadır. Şekil 9 ve Şekil 10'daki karşılaştırmalardan elde edilen sonuçlara göre Rastgele Orman algoritması diğer algoritmalara göre daha iyi sonuçlar vermiştir. Buna göre Rastgele Orman algoritması bu çalışmada kurulacak model için kullanılmıştır.



**Şekil 11.** Hindistan bölgesinde ilk 6 ay için yetiştirilebilecek ürün harita (Map of crops that can be grown in India for the first 6 months)



**Şekil 12.** Hindistan bölgesinde ikinci 6 ay için yetiştirilebilecek ürün haritası (Map of crops that can be grown in India for the second 6 months)

Rastgele Orman algoritması kullanılarak Hindistan tarımsal ürün haritası yılın ilk 6 ayı ile ikinci 6 ayı için ayrı ayrı yapılmıştır. Yılda birden fazla ürün alınacağı düşünülmüştür. Hindistan bölgesinde ilk 6 ay için yetiştirilebilecek ürün haritası Şekil 11 'de gösterilmiştir. Yine aynı şekilde Hindistan bölgesinde ikinci 6 ay için yetiştirilebilecek ürün haritası Şekil 12 'de gösterilmiştir. Şekil 11 ve 12'de gösterildiği gibi farklı tarım ürünleri farklı renklerle ifade edilmiştir. Hindistan'ın bazı bölgelerinde aynı ürün yılın ilk 6 ayı ve ikinci 6 ayı içinde yetiştirilebilecekken bazı bölgelerde yılın ilk 6 ayı ve ikinci 6 ayı içinde farklı ürünler yetiştirilebilmektedir.

### 5. Sonuç ve Değerlendirme (Result and Evaluation)

Literatürdeki benzer çalışmalar ile bu makalede önerilen modelin hangi akıllı tarım uygulamalarına hangi yöntemlerin uygulandığı ve uygulanan yöntemlerden en başarımı yüksek model ve başarımların değeri Tablo 1'de gösterilmiştir. Literatürde farklı akıllı tarım uygulamalarına farklı yöntemler uygulanmaktadır. Tablo 1'de de görüldüğü gibi bu makalede geliştirilen modelin doğruluk değeri diğer çalışmalara göre daha yüksektir. Ayrıca, bu makale literatürdeki diğer çalışmalardan farklı olarak sistem başarımında modelin çalışma zamanını da değerlendirmeye almaktadır. Akıllı tarım uygulamalarında veri boyutu gittikçe artmaktadır. Bu nedenle algoritmanın çalışma zamanı bazı durumlarda algoritmanın doğruluk değerinden daha fazla dikkate alınmaktadır.

**Tablo 1.** İncelenen çalışmalar ve önerilen çalışmanın karşılaştırmalı değerleri (Comparative values of the studies reviewed and the proposed study)

Kaynak	Akıllı Tarım Uygulaması	Kullanılan Yöntemler	En Başarılı Yöntem	Sistem Başarısı
(Muangprathuba vd., 2018)	Akıllı Sulama Sistemi	Birliktelik Kuralları (Association rules) (Apriori algoritması), Doğrusal regresyon	Birliktelik kuralları (Apriori algoritması)	-
(Rodríguez vd., 2021)	Kahve Üretim Tahmini	Karar Ağaçları, Yapay Sinir Ağları, XGBoost, Destek Vektör Makinesi, Rastgele Orman	XGBoost Algoritması	Ortalama Mutlak Hata:0.032
(Sujatha vd., 2021)	Görüntü Sınıflandırma ve Görüntü Algılama ile Turunçgiller Bitkisi Hastalık Tespiti	Destek Vektör Makinesi, Rastgele Orman, Olasılıksal Dereceli Azalma, Derin Öğrenme (Inception-v3, VGG-16, VGG)	VGG-16	Doğruluk: %89,5
(Horng vd., 2019)	Akıllı Görüntü Tanımaya Dayalı Bir Hasat Tespiti Sistemi	Inception, ResNet, MobileNet, Faster R CNN, YOLO, SSD,	MobileNet-SSD	Ortalama Kesinlik Değerlerinin Ortalaması %84
(Wang vd., 2021)	Şeker Üretim Miktarının Tahmini	Çok Katmanlı Algılayıcı Sinir Ağı, Destek Vektör Makinesi, Rastgele Orman, Doğrusal Regresyon, Karar Ağacı, Naive Bayes	Çok Katmanlı Algılayıcı Sinir Ağı	Doğruluk: %99
Yapılan Çalışma	Tarım Ürünü Haritası	K- En Yakın Komşu, XGBoost Algoritması, Rastgele Orman	Rastgele Orman	Doğruluk: %99,5 Çalışma zamanı: yaklaşık 0,05 s

Bu çalışmada Akıllı tarım sistemlerinde incelenen makalelerde, hasat tahmini, yabancı ot tespiti, sulama, tarımsal zararlıların ilaçlanması vb. gibi birçok problem, insansız hava araçları (İHA) ve robotlar, IoT cihazları, yapay zekâ (AI), Coğrafi Bilgi Sistemleri (CBS) derin öğrenme (DL), makine öğrenimi (ML) ve kablosuz iletişim kullanılarak gerçek zamanlı olarak ele alınmıştır. Bu çalışma Akıllı tarımda veri analizi ile ürün tahmini üzerine odaklanmıştır. Kaggle üzerinden sağlanan toprak verileri (ph değeri, azot değeri, potasyum değeri ve fosfor değeri) iklim verileri (sıcaklık yağış ve nem) kullanılarak veri madenciliği algoritmaları ile farklı modeller oluşturulmuştur. Oluşturulan modeller doğruluk, kesinlik, duyarlılık, f-skör ve algoritmanın çalışma zamanına göre kıyaslanmıştır. Elde edilen sonuçlara göre doğruluk, kesinlik, duyarlılık ve f-skör değerlerine göre en kötü sonuçları KNN en iyi sonuçları Rastgele Orman algoritması vermiştir. Modeller çalışma zamanlarına göre kıyaslandığında XGBOOST algoritması en kötü sonucu vermiş, Rastgele Orman algoritması en iyi sonucu vermiştir. Her iki karşılaştırmada da en iyi sonucu veren Rastgele Orman ile kurulan model, Hindistan bakanlıklarınca sağlanan verilere uygulanmış ve Hindistan'ın ilk altı ay ve ikinci altı aylık tarımsal ürün haritası oluşturulmuştur. Tarımsal ürün haritasına göre belirli bölgelerde yılın ilk altı ayı ve ikinci altı ayı için yetiştirilebilecek ürünler değişmezken belirli bölgelerde yılın farklı aylarında farklı ürünler yetiştirilebilmektedir.

Günümüzde iklim konusu önemli bir konuma sahiptir. Günümüzde belli bölgeler iklim ve toprak özelliklerine göre belirli tarım ürünleri ile özdeşleşmiştir. İklim özelliklerinin hızla değiştiği bu bölgelerde yetişen ürünlerden ileride yeterince verim alınamayacaktır. Bu bölgeler artık farklı tarım ürünleri için daha uygun olacaktır. Bazı tarım ürünleri çok yıllık olup yetiştirilmesi ve ürün vermesi yıllar almaktadır. İleriye yönelik iklim özellikleri tahminlerinin gelişmesi, şimdiden bazı tedbirlerin alınmasını gerektirmektedir. Bu çalışma ile ileriye dönük tahmin edilen iklim verileri ile bölgelerin gelecek yıllar için ürün haritası oluşturulabilmektedir.

### Çıkar Çatışması (Conflict of Interest)

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir. No conflict of interest was declared by the authors.

### Kaynaklar (References)

- Annual Rainfall Map of India (n.d.). <https://www.mapsofindia.com/maps/india/annualrainfall.htm>  
Average Humidity for India in January (n.d.). <https://www.currentresults.com/Weather/India/humidity-january.php>  
Balducci, F., Impedovo, D., & Pirlo, G. (2018). Machine Learning Applications on Agricultural Datasets for Smart Farm Enhancement. MDPI, machines, 6(38), 1-22.



- CLIMATE-SMART AGRICULTURE. (2021, Nisan 5). <https://www.worldbank.org/en/topic/climate-smart-agriculture>.
- Ensemble methods (2022). <https://scikit-learn.org/stable/modules/ensemble.html#forest>
- Horng, G.-J., Liu, M.-X., & Chen, C.-C. (2019). The Smart Image Recognition Mechanism for Crop Harvesting System in Intelligent Agriculture. *IEEE Sensors Journal*, 1-16.
- Idoje, G., Dagiuklas, T., & Iqbal, M. (2021). Survey for smart farming technologies: Challenges and issues. *Computers & Electrical Engineering*, 96, 1-14.
- India Current Temperature [°C] (2021). <https://www.weatheronline.in/weather/maps/current?LANG=in&DATE=1604127600&CONT=inin&LAND=II&KEY=II&SORT=2&UD=0&INT=06&TYP=temperatur&ART=karte&RUBRIK=akt&R=310&CEL=C&SI=kph>
- Ingle, A (2020, Aralık 2020). Crop Recommendation Dataset. [https://www.kaggle.com/atharvaingle/crop-recommendation-dataset?select=Crop\\_recommendation.csv](https://www.kaggle.com/atharvaingle/crop-recommendation-dataset?select=Crop_recommendation.csv)
- Li, W., Zheng, T., Yang, Z., Li, M., Sun, C., & Yang, X. (2021). Classification and detection of insects from field images using deep learning for smart pest management: A systematic review. *Ecological Informatics*, 66(101460), 1-18.
- Mohamed, E. S., AA.Belal, Abd-Elmabod, S., El-Shirbeny, M. A., A.Gad, & Zahran, M. B. (2021). Smart farming for improving agricultural management. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3), 971-981.
- Muangprathuba, J., Boonnama, N., Kajornkasirata, S., Lekbangponga, N., Wanichsombata, A., & Nillaorb, P. (2018). IoT and agriculture data analysis for smart farm. *Computers and Electronics in Agriculture*, 467-474.
- Nearest Neighbors (n.d.). <https://scikit-learn.org/stable/modules/neighbors.html#classification>
- Ok, A., Akar, Ö., & Gungor, O. (2011). Rastgele Orman Sınıflandırma Yöntemi Yardımıyla Tarım Alanlarındaki ürün Çeşitliliğinin Sınıflandırılması. *TUFUAB 2011 VI. Teknik Sempozyumu*, (s. 1-7). Antalya.
- Pathak, A., AmazUddin, M., Abedin, M. J., Andersson, K., Mustafa, R., & Hossain, M. S. (2019). IoT based Smart System to Support Agricultural Parameters: A Case Study. *Procedia Computer Science*, 155, 648-653.
- Podder, A. K., Bukhari, A. A., Islam, S., Mia, S., Mohammed, M. A., Kumar, N. M., . . . Abdulkareem, K. H. (2021). IoT based smart agrotech system for verification of Urban farming parameters. *Microprocessors and Microsystems*, 82(104025), 1-10.
- Ratnaparkhi, S., Khan, S., Arya, C., Khapre, S., Singh, P., Diwakar, M., & Shankar, A. (2020). Smart agriculture sensors in IOT: A review. *Materials Today: Proceedings*, 1-6.
- Rodríguez, J. P., Montoya-Munoz, A. I., Rodríguez-Pabon, C., Hoyos, J., & Corrales, J. C. (2021). IoT-Agro: A smart farming system to Colombian coffee farms. *Computers and Electronics in Agriculture*, 190, 1-18.
- Roukha, A., Fotea, F. N., Mahmoudia, S. A., & Mahmoudia, S. (2020). Big Data Processing Architecture for Smart Farming. *The 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks*. Madeira.
- Soil Maps- Cycle I (n.d.). <https://soilhealth.dac.gov.in/NewHomePage/SoilMap>
- South Asia Network on Dams, Rivers and People. (n.d.). <https://sandrp.in/category/rainfall/page/2/>
- Sujatha, R., Chatterjee, J. M., Jhanjhi, N., & Brohi, S. N. (2021). Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocessors and Microsystems*, 80, 1-11.
- Tay, B., Hyun, J. K., & Oh, S. (2014). A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained from Diffusion Tensor Images. *Comput Math Methods Med*.
- Wang, P., Hafshejani, B. A., & Wang, D. (2021). An improved multilayer perceptron approach for detecting sugarcane yield production in IoT based smart agriculture. *Microprocessors and Microsystems*, 82(103822), 1-7.
- XGBoost Documentation. (n.d.). <https://xgboost.readthedocs.io/en/stable/>